National Compliance Assistance Providers Forum
"Optimizing Resources for Environmental Results"

December 3-6, 2002
San Antonio, Texas

The following are annotated meeting notes recorded by Rick Yoder
(ryoder@unomaha.edu) and transcribed by Scott Butner (scott.butner@pnl.gov) from the
session "500 Channels and Nothing's On," which took place at the 2002 Compliance
Assistance Provider's Forum, in San Antonio, Texas.

These notes are transcribed in approximate order of the issues/points identified, but in
some cases they have been re-organized to bring recurring themes together into a single
topic heading.

Comments, questions or concerns about these notes can be directed to
scott.butner@pnl.gov or Paul Chalmer paul.chalmer@ncms.org, (734) 995-4911.

Participants expressed a need for an email list to discuss EPA web issues from a user's
perspective rather than a developer's perspective.   We are polling the community to see
if a list is needed and/or if one exists already to meet this need.   Please contact Scott,
Paul, or Greg Geyer greg@terrachord.com if you are interested in this list.

"What Works? What Doesn't?"

The "web scavenger hunt" questions (with answers) posted by Paul Chalmer in the
registration material prior to the conference, are available at the following URLs:

http://ecm.ncms.org/hunt/index.html
http://ecm.ncms.org/hunt/answers.htm

The scavenger hunt exercise provided good insight into the difficulties facing web
researchers who attempt to locate online compliance assistance information, which often
concerns itself with very precise meanings, and requires attention to authenticity and
authority of the data sources. These challenges led to a broader, free-form discussion of
what works and what doesn't vis-à-vis finding information on the web.

Google works well, is becoming the defacto search engine standard. But it works best
when tuned with the right search terms.
- Some concerns about what to do when Google comes up short
- Reminder that different search engines often produce different results -- this
  highlights the potential danger of monopoly by any one engine.
- Meta-search tools (dogpile, etc.) were mentioned as useful alternatives to the "one
  size search engine fits all" problem. Meta-search tools compile the results from

several search engines, based on a single search. In some cases the results are kept separate, but in other cases, they are combined into a single search result set. Several examples are listed below.
- o Meta-search alternatives to Google
- o Dogpile (http://www.dogpile.com/)
- o Metacrawler http://www.metacrawler.com/index.html
- o Ixquick (http://ixquick.com/eng/aboutixquick/)

Some other approaches to searching which are in developmental stages at Google Labs:
- http://labs.google.com/cgi-bin/webquotes
  This site extracts site summaries from other sites, so you can find out what others are saying about your site (or any other site!)
- http://labs.google.com/sets
  This site lets you enter a number of terms, and finds others that correlate.   It's cooler than it sounds.
- http://www.touchgraph.com/TGGoogleBrowser.php?start=www.p2rx.org
  The last one does some very interesting things with web connectivity, showing how things are "wired together" on the web.

"Collaborative filtering" is the use of a community of like-minded people to help add value to information. Perhaps the most well known examples are user reviews on Amazon.com, or the "people who bought this book also bought…." feature on the same web site.
- Google also uses a form of collaborative filtering, by factoring how many other sites link to a given site as part of the relevance ranking. This is an indirect measure of how popular a site is within a community of interest.
- There is power in collaborative filtering (technique used by Google, Amazon, etc.) which could be better tapped within the technical assistance community.
- The real challenge is determining what are the information "gems" – key or "core" documents which represent the "best of the best"
- P2Rx (www.p2rx.org) has tried to capture some of this via their pollution prevention "topic hubs" (accessible through their web site).

What are needed are practical solutions to finding information in a more selective, focused, targeted manner.

How to Find Information?

There was some discussion of search strategies, which came about because several people in the audience expressed an interest in "tips from experts" on finding information.
Search "tips"
- Craft search string
- Form a precise question (develop a clear mental picture of what it is you are looking for, so you can recognize it when you see it)

- Take advantage of subject-specific acronyms, "buzz phrases" that you can use to target the intended subject matter
- Use familiar sources as starting points
- Use the search engine to find good "jumping off points" which have highly focused links to related subject matter
- Take advantage of advanced search options
- Use quotes to enclose subject specific phrases
- Consider using the Google toolbar (http://toolbar.google.com/) to help make searching more effective. The Google toolbar adds new capabilities to your browser, such as highlighting of search terms in the results, etc.
- Context-aware search engines offer some hope for making searching more effective
- Example given was CiteSeer http://citeseer.nj.nec.com/cs which provides a wide variety of context clues about searching for papers in the Computer Science literature. These include an analysis of how often each paper is cited in other publications.
- Ask Jeeves (http://www.ask.com/) can be an effective search tool, but only for certain "types" of questions. For example, it does a good job of returning answers to commonly asked questions in broad subject domains (e.g., "how long should I cook my turkey?"). It is less effective for specialized and technical topics (e.g., "Where can I find the pulp and paper MACT?") though will often return useful search results nonetheless.

Discussion on EPA's Web Site

The discussion moved to specific issues related to EPA's information and the organization of EPA web space. Some highlights (?) below:

General Observations

- Regs can be relatively easy to find
- Other information is often very hard to find
- EPA search engine returns too many internal documents (memos, guidance docs, etc)
- "EPA Topics" help organize the site into broad subject areas
- Recent navigational changes have, to some extent, helped
- Conformity of interface makes it easier to locate info on the page
- Content tends to focus on public access/right-to-know, rather than compliance assistance/pollution prevention
- Different Audiences & Needs
  - EPA information accessibility depends a lot on organizational perspective
  - Industry seeks info from customer perspective
  - EPA information deployment retains credibility

Issues Limiting Effectiveness

- EPA faces many institutional barriers that are often echoed in other agencies
- "Conservative" IT policies
- Need to balance site security (e.g., protection against hackers, protection of Confidential Business Information which EPA has access to) against interactivity, openness.
- Congressional oversight and occasional intervention
- Plethora of largely autonomous regions, program offices, initiatives creating their own sites
- Putt's law applies: (http://www.onfocus.com/quote.asp?author=265) : "Technology is dominated by two types of people: those who understand what they do not manage, and those who manage what they do not understand"

Tips on searching using EPA search engine

- Quote document title if you know it

"Home Brewed"Tips

- Personal filing system (e.g., in Internet Favorites or bookmarks file) can allow easy classification of useful documents
- Potential problem is the difficulty in maintaining the resource once you accumulate many URL's.
- Link Doctor (http://www.envcap.org/ld/) helps to maintain such lists.   The site was developed/operated by NCMS as part of their Center Platform project, and maintains a history of outdated URLs so you can search for the new location of a "missing in action" document.

How to Help Industry?

Solution (?): community-built knowledge bases to retain institutional memory

Stellar example: P2TECH (http://www.great-lakes.net/lists/p2tech/index.html)

- Gap between regional and national efforts
- EPA web group and templates
- EPA list serves are great resources
- Industry Info Needs
- Web may not be best resource for all users
- Trade organizations haven't been pulled into Clearinghouse (http://www.epa.gov/clearinghouse/)
- Delivery route: first get Technical Assistance attention, then direct to web for follow-up resources

- Needs interactivity (e.g., OSHA compliance wizards)

What's Next?

Possible startup of a list server to discuss EPA web issues from a user's perspective rather than a developer's perspective.

Miscellaneous Items

May 2001 "Scientific American" article by Tim Berners-Lee, James Hendler and Ora Lassila on "The Semantic Web."
http://www.sciam.com/article.cfm?articleID=00048144-10D2-1C70-84A9809EC588EF21

According to one panelist (Butner) the semantic web concepts espoused by Berners-Lee, et al, will transform the web as we know it, probably in the next 3-5 years. The emphasis will shift towards how documents are related (to each other, to other documents, to concepts and issues) rather than which words they contain. This article is a very readable view of how this will happen.

"TOP 4 THINGS THE WEB NEEDS" List by Scott Butner
Interactivity. Users expect to interact with information in a more active way, even if this doesn't always mean discussion lists, etc. Simply giving people some way to influence what's on the screen is important. Expect to see information become more interactive. E.g., OSHA's "regulatory wizards" (for example, see http://www.dol.gov/elaws/confined.htm). Even simpler example might be: http://toxtown.nlm.nih.gov/

Interoperability. Our other applications all interact with each other – we can export data from Word to spreadsheets, to databases, etc. Why not the web?   Expect to see more options for exporting data from web sites into other applications. Example: Compliance Assistance Clearinghouse can return search results in the form of a file which can be imported into databases, spreadsheets.
http://cfpub.epa.gov/clearinghouse/webservices/about_remote_search.cfm

Context. Information is useful in context. Expect to see ideas from the Semantic Web infiltrate the web and begin allowing authors to put their documents in context.

Community. See discussion above about P2TECH and collaborative filtering.